Flinders University
2024 S1 COMP2031 Data Engineering
Project Description + Marking Criteria

Overview:

You are required to work in a group working environment of five students with the below-selected data collection from MongoDB's sample data collection "grades" and the Open University Learning Analytics (OULAD) dataset. You can select to work either on "grades" data collection or OULAD dataset; or you can work on both grades" data collection and OULAD dataset. You will use R code that can run on RStudio, including all required packages. You will be graded on the quality of your data transformation and analysis, the accuracy and effectiveness of your data models and visualisations, and the clarity and persuasiveness of your report and presentation.

Although this is a group working environment, the project is an individual assessment. It means that each group member will be assessed based on the individual work, as well as the individual submission and individual performance during the presentation. However, within the group working environment, group members are expected to actively support other group members if it is applicable.

Selected data collection at MongoDB sample dataset:

MongoDB's sample datasets	Data collections
sample_training	grades

Descriptions of each dataset can be found at the following link:
https://www.mongodb.com/docs/atlas/sample-data/

Open University Learning Analytics (OULAD) dataset:

Descriptions of the dataset or how to download the data set can be found at the following link

https://analyse.kmi.open.ac.uk/open_dataset

Project requirements and general marking criteria:

The project requirements are listed in the following tasks:

1) Data Wrangling: Loading and tidying the dataset to ensure it is in a clean and usable format.
   a) Loading the data accurately
   b) Handling missing or incorrect data appropriately
   c) Tidying the data effectively
2) Data Transformation: Using data transformation techniques to transform the dataset into a format suitable for analysis.
   a) Using appropriate data transformation techniques to prepare the data for analysis.

b) Creating new variables as needed
3) Data Analysis: Analyse the dataset using appropriate statistical methods and create data visualizations to support your analysis.
   a) Using appropriate statistical methods to analyse the data
   b) Creating accurate and effective data visualizations
4) Data Modelling: Creating data models for the relevant variables in the dataset.
   a) Creating a simple linear model
   b) Creating a general linear model. The model should cover three types of predictors which are
      i) Predictors are categorical.
      ii) Predictors are categorical and continuous.
      iii) Predictors are continuous.
   c) Evaluating the performance of the models if applicable
   d) Interpretability of the models

Additional marking criteria for project report:

Task 5 of the project report: The report should cover a clear and concise summarizing your findings, including relevant visualizations and models. The report should not exceed 10 pages (excluding title page, table of contents, references and appendices). The word count limit is 3000 words, (excluding figures, tables, title page, table of contents, references and appendices). Note that the title page, table of contents, references and appendices are not within the scope of marking.

Task 6 of the project report: A Group Support Statement should cover details of all members' support to other members. Each team member is required to submit the group support statements weekly and separately.

Tasks	Report (50 marks)
1	20%
2	20%
3	20%
4	20%
5 : Report presentation	10%
6 : Group Support Statement	10%

Additional marking criteria for project oral presentation:

Task 5 of the project presentation: The slides and presentation should provide an engaging and informative presentation of your findings. The presentation should be no longer than 5 minutes for student.

Tasks	Presentation (20 marks)
1	20%
2	20%
3	20%
4	20%
5 : Slides and presentation	20%

Additional marking criteria for project demo oral presentation:

Task 5 of the project demo: You are required to submit an R source code file along with any additional files, packages or data needed to run the code. The source code file must be runnable on R studio and include all necessary code, and comments to allow the markers to reproduce the tasks. The R code should provide appropriate annotations for the associated tasks, reference points to the pages of the report and pages of the presentation slides. The code demonstration should provide the demonstration that the R code can run on RStudio and cover the tasks.

Tasks	Demo (10 marks)
1	20%
2	20%
3	20%
4	20%
5 : Code presentation & demonstration	20%

Group registrations and group presentation modes:

Students are expected to form groups within their registered tutorial class and students are responsible for forming their groups. Group registration will be organised at the tutorial classes. Each group will present either online or on-campus depending on the class they are enrolled in. Groups of the COMP2031 online class will present their work online, while groups of the COMP2031 on-campus class and the COMP8031 on-campus class will present their work on-campus. Each group is required to register for a 30-minute oral presentation slot. Within the presentation slot, each team member has a 5-minute oral presentation limit to cover the project oral presentation and project demo presentation.

Plagiarism:

Plagiarism, which includes copying from internet resources or other group members, is not acceptable and will result in a grade of zero for the entire project. While using external sources for reference and inspiration is allowed, all work submitted must be original and produced by each individual. Any external sources used must be appropriately cited and referenced in the project report. You must understand the requirements for academic integrity to ensure that your submissions are original and produced by yourself.

Using identical demonstrations or operations on the same data as other groups is also not allowed. Each member is expected to perform their own unique analysis and modelling on their chosen data. While there may be some overlap in the techniques used, the overall approach and conclusions drawn should be distinct for each member. Any evidence of copying or using identical examples from other students or external sources will result in a grade of zero for the entire project.

You are not allowed to use identical examples or demonstrations on the same data from the teaching materials in the workshops and tutorials. While these resources are provided to aid learning and understanding, the purpose of the project is to assess your ability to independently apply the concepts and techniques covered in class.